VARIANCE ESTIMATION IN NEW RESIDENTIAL CONSTRUCTION SURVEY

John Aleong, Diantha B. Howard, and Thomas Patterson, U. of Vermont
John Aleong, U. of Vermont, Burlington, VT 05405

ABSTRACT

A survey was conducted to determine new construction practice, particularly related to energy conservation, in the state of Vermont during 1986. The sampling frame was constructed from the Bureau of Census Construction Report C40-85-13 for the State of Vermont (1985). Within eight counties, towns were stratified into certainty and noncertainty towns, based on the census definition. Within each stratum, clusters were formed of one town each for the certainty towns, and groups of neighboring towns for the noncertainty towns. Within a stratum, clusters were sampled with replacement, by probability proportional to their size (the number of building permits issued). Within a cluster, a systematic sample without replacement of permits was taken. In order to obtain estimated standard errors for population estimates from the survey, we used the sampling program TREES, developed by Bellhouse and Rylett. This program calculates estimates of the variance-covariance matrix for estimates of means, totals, and proportions at any stage of a multi-stage sampling design.

INTRODUCTION

As partial fulfillment of a contract for FY87 with the Vermont Department of Public Service, the University of Vermont Extension Service Energy Program conducted a sample survey of 152 new residential structures to determine current energy efficient construction practices. The survey was conducted from August through November 1986.

A survey of new residential construction had never been attempted in Vermont. Policymakers, government officials, and the Vermont construction industry were using estimates based, in the best case, on incomplete statistical information, and, in the worst case, on speculation.

A survey of new building construction would provide a data base for:

a. The construction industry in determining current construction practices and in improving the energy efficiency of new buildings.

b. State government agencies in planning for future energy and housing needs.

c. Policymakers in developing energy and building policies for the State of Vermont.

METHODS

1. Sample Selection

Eight counties were selected which were within reasonable driving distances for the Extension Service energy agents, who were the interviewers. 120 single family units and 32 multi-family units were surveyed.

The sampling frame was constructed using similar methodology to that used by the Bureau of Census for the Construction Report C40 series. Within each county, towns were stratified into certainty and noncertainty towns, according to the Census definition for the 1983 Building Permit Survey.[1] Within each stratum, clusters were formed of one town each for the certainty towns, and groups of neighboring towns for the noncertainty towns. Within a stratum, clusters were sampled with replacement, by probability proportional to their size (the number of building permits issued in 1985, reported in Construction Report C40-85-13). This was done for single family and multi-family units separately. One town was sampled from each of the noncertainty clusters, again with probability proportional to size. For each town, (certainty and noncertainty), a systematic random sample of permits was obtained.

For the state of Vermont, there were 82 certainty and 93 noncertainty places. The eight selected counties comprised 62 certainty and 63 noncertainty places. The 1985 Construction Report showed 1746 permits issued for these counties. The sample of 120 single family and 32 multi-family permits was allocated proportional to the size of the certain and noncertain strata for each county. Within each stratum, the sample was allocated among the towns, again proportional to their size. For each selected town, the agents recorded the total number of building permits (both single and multiple) issued from January 1, 1986 to the time the sampling took place. From this total they took a systematic sample without replacement to select the specified number of permits. If the chosen site was not suitable for the survey, the agent would revert to a backup site also selected at this time. Utilizing building permits, there was no way to determine the stage of construction of the selected building site, so visual inspection was necessary. The agents found sites ranging from an empty lot to an occupied house.

2. Questionnaire

A literature search and a review of UVM Extension Service and Department of Public Service (DPS) files failed to uncover a questionnaire suitable for a survey of new housing construction practices, focusing on energy efficiency. A draft questionnaire was drawn up and circulated for feedback among energy agents and representatives of the DPS. Their input was incorporated into the final 11 page questionnaire.

The questionnaire addressed the following areas of inquiry: general information, building construction of foundations, walls, ceilings and attics, ventilation, and installed electrical load. In addition to construction materials and practices, workmanship issues were also addressed. This workmanship category also enabled the agents to indicate whether or not they had actually observed the particular practice addressed in the questionnaire.

Definitions of terms found in the questionnaire were established by consensus at a training session held for the four energy agents and a representative of the DPS on 24 July 1986. The questionnaire was finalized after a pilot testing at several new single family residence sites.

3. Estimation

For the analysis of the survey, the following sample design was assumed for convenience. Certainty and noncertainty strata were combined within a county. Then within a county, clusters were sampled with replacement with varying probabilities, and a simple random sample of permits was drawn without replacement from each cluster. The certainty and noncertainty strata were combined since unbiased estimates of variance could not be computed with only one psu drawn from a stratum (Sukhatme and Sukhatme, 1970, p. 334). This occured for all the noncertainty strata. Also, the subsampling of a town within the nondertainty clusters was ignored. Note that the use of a simple random sample for the subsamples, rather than a systematic sample, may give approximate variances.

Let

k = number of strata

N_t = number of psu in the t-th stratum

P_{ti} = selection probability of the i-th psu from the t-th stratum, where

$$\sum_{i=1}^{N_t} P_{ti} = 1 \text{ for each } t = 1, \ldots k$$

M_{ti} = number of ssu in the i-th psu from the t-th stratum

M_{t0} = total number of ssu in the t-th stratum

$\alpha_t = M_{ti}/M_0$, where M_0 = number of ssu in population

y_{tij} = value of the j-th ssu from the i-th psu in the t-th stratum

n_t = number of psu sampled from the t-th stratum

m_{ti} = number of ssu to be selected from the i-th psu in the t-th stratum, if it is in the sample

τ_{ti} = number of times the i-th psu of the t-th stratum selected, such that

$$\sum_{i=1}^{N_t} \tau_{ti} = n, \text{ for each } t.$$

Let $z_{tij} = M_{ti}y_{tij}/(M_{t0}P_{ti})$.

An unbiased estimator of the population mean is

$$\bar{z} = \sum_{t=1}^{k} \alpha_t \bar{z}_{ts} \text{ , where}$$

$$\bar{z}_{ts} = (1/n_t) \sum_{i=1}^{n_t} \tau_t \bar{z}_{ti} \text{ , and}$$

$$\bar{z}_{ti} = (1/m_{ti}\tau_{ti}) \sum_{j=1}^{m_{ti}\tau_{ti}} z_{tij}.$$

An unbiased estimator of the variance of $\bar{z}$ is

$$\hat{V}(\bar{z}) = \sum_{t=1}^{k} \alpha_t^2 \hat{V}(\bar{z}_{ts}), \text{ where}$$

$$\hat{V}(\bar{z}_{ts}) = (s_{tbz}^2/n_t) + [1/(n_t(n_t-1))] \sum_{i=1}^{n_t} [1/m_{ti} - 1/(m_{ti}\tau_{ti})] s_{tiz}^2 - (1/n_t) \sum_{i=1}^{n_t} P_{ti}s_{tiz}^2/M_{ti},$$

(Sukhatme & Sukhatme, eq. 101, expanded for stratified sampling), where

$$s_{tbz}^2 = [1/(n_t-1)] \sum_{i=1}^{n_t} \tau_{ti} (\bar{z}_{ti} - \bar{z}_{ts})^2,$$

and

$$s_{tiz}^2 = [1/(m_{ti}\tau_{ti}-1)] \sum_{j=1}^{m_{ti}\tau_{ti}} (z_{tij} - \bar{z}_{ti})^2.$$

4. Software

The computer program TREES was developed by David R. Bellhouse and David

T. Rylett to produce population estimates for complex multi-stage survey designs (Bellhouse, 1980 and 1985). The program calculates estimates of the variance-covariance matrix for estimates of means and totals. It uses the general formula for estimating the variance in multistage designs derived by Rao (1975). The program currently allows three unistage sampling designs: simple random sampling without replacement and two methods of sampling without replacement, with probability proportional to size. A routine was added by one of the co-authors to handle simple random sampling with replacement and probability proportional to size.

RESULTS

This paper gives means and variances for the most important items in the questionnaire for single family units only. Several of the variables of interest were dichotomous, which were coded 1-0. For the polychotomous variables it was necessary to create new 1-0 variables for each category, and calculate means and estimated variances for each category separately. Results are shown in Table 1.

---

1. Personal communication, Donald M. Luery, Chief Research and Methods staff, Construction Statistics Division, U.S. Department of Commerce Bureau of the Census.

All places which authorized housing units during 1978, 1981, and 1982 greater than a predetermined number of permits for either residential or nonresidential projects were certainty places.

FURTHER WORK

More work needs to be done on estimating the variances of categorical variables and expanding the TREES program for other sampling designs.

REFERENCES

Bellhouse, D.R., (1980), "Computation of variance-covariance estimates for general multistage sampling designs," COMPSTAT 1980: Proceedings in Computation Statistics, pp. 57-63, Physica-Verlag, Vienna.

Bellhouse, D.R., (1985), "Computing methods for variance estimation in complex surveys," Journal of Official Statistics, 1(3):323-329.

Rao, J.N.K., (1975), "Unbiased variance estimation for multistage designs," Sankhya, series C, 37:133-139.

Sukhatme, P. V. and Sukhatme, B. V., (1970), Sampling Theory of Surveys with Applications, Iowa State University Press, Ames, Iowa.

Table 1.

Selected Energy-related items from Questionnaire

Item from Questionnaire	Mean (%)	Estimated St. Dev.
1. Covered by Act 250	29.6 %	5.8
2. Primary Residence	92.4 %	3.5
5. Constructed under Energy Covenants	25.0 %	5.7
7. Approximate floor area (excluding basement):	1853 sq. ft.	77.8
11. Type of construction:		
Stick-built:	90.9 %	3.1
Log:	2.6 %	1.6
Modular:	1.8 %	1.4
Pre-fabricated:	0.5 %	.5
Other:	4.1 %	2.7
14. Foundation insulation material:		
Extruded polystyrene:	46.5 %	6.3
Fiberglass:	7.4 %	3.7
Expanded polystyrene:	0.9 %	.9
Other:	1.2 %	.8
No foundation insulation:	44.0 %	7.7
22. Vapor barrier:		
Film:	62.9 %	8.0
Integral with insulation:	26.5 %	8.2
Aluminum-faced dry wall:	1.4 %	1.3
No vapor barrier:	9.3 %	2.5
25. Wall Construction:		
2x4 wood stud:	6.8 %	1.4
2x4 metal stud:	0.9 %	.9
2x4 double stud, 8 in. wall:	1.7 %	1.7
2x6 wood stud:	81.9 %	3.7
Post and beam:	5.3 %	2.8
Logs:	2.6 %	1.6
Other:	0.7 %	.7
65. Primary source of heat:		
Oil:	49.0 %	7.0
Gas:	27.9 %	6.6
Wood, Coal:	14.3 %	3.8
Electric, uncontrolled:	4.7 %	2.1
Other:	3.8 %	2.0
79. Sited to take advantage of incidental solar radiation	40.6 %	5.4
80. Obvious solar design	16.0 %	4.1
81. Annual energy cost calcuated	9.4 %	2.4
83. Surveyor's overall rating of unit's energy efficiency:		
Excellent:	5.4 %	3.4
Good:	49.5 %	6.9
Fair:	37.3 %	6.1
Poor:	7.8 %	2.4